Eng Int Syst (2024) 6: 647–657

Engineering
Intelligent Systems

English Learning Behavior Pattern Mining and Personalized Teaching Strategies Based on Big Data Analysis

Yingying Xiao*

School of Education, Jingzhou University, Jingzhou 434000, Hubei, China

With the development of information technology, data-driven decision-making in education is becoming increasingly important, especially in the personalization of English language teaching. In this study, large-scale English learning behavior data were deeply mined through a set of well-established analysis processes, using quantitative methods such as cluster analysis and association rule analysis. It was found that the careful delineation and parsing of students' behavioral patterns revealed individual differences in terms of learning habits, preferences, and challenges faced by students when learning English. The experimental results show that the experimental group that implemented personalized teaching strategies demonstrated more significant improvement in the learning behaviors (e.g., online learning hours, number of interactions, task completion, etc.) and academic performance (e.g., test scores, homework grading) than those of the control group that were exposed traditional teaching modes. The case analyses of specific cases also demonstrated that the personalized teaching strategy designed according to students' individual characteristics can effectively improve the learning outcomes of students.

Keywords: big data analytics, English learning, behavioral pattern mining, personalized instruction

1. INTRODUCTION

Under the wave of global informatization, the education sector is undergoing a profound transformation and innovation, in which the introduction and development of big data technology undoubtedly plays a key role in driving the wave. With its powerful data collection, storage, processing and analysis capabilities, big data technology revolutionizes the traditional education model and opens a new chapter of intelligent educational decision-making and teaching activities. Especially in the field of English teaching, the application of big data technology enables educators to pay attention to the learning dynamics of each student in a comprehensive and multi-dimensional way, and capture their subtle changes and deep-rooted needs in the learning process [1–3].

*Corresponding author's Email: xiaoyingying888@hotmail.com

By means of big data analysis, we can identify patterns in students' learning behaviors, as well as their academic interests, learning paths, learning difficulties, etc., so as to obtain a more accurate and thorough understanding of the English learning process. However, it is often difficult for the existing English teaching model to meet the different individual learning needs of students, although the refined and personalized teaching approach facilitated by big data analysis makes this possible. By conducting an in-depth analysis of learning behavior patterns, educators can formulate appropriate teaching strategies according to the characteristics of each student, such as targeted knowledge push, personalized curriculum, timely learning feedback, etc., so as to achieve the essence of tailor-made teaching and improve the efficiency of teaching and learning effectiveness [4–6].

This study uses big data analysis technology to explore the behavioral patterns of English language learners in order to identify the intrinsic connections and patterns of their learning habits, learning efficiency, learning difficulties and other elements of learning. Then, according to the characteristics of these behavioral patterns [7], we can scientifically construct and optimize personalized English teaching strategies, realize tailor-made teaching, improve teaching quality and effect, and bring the potential of each student into full play. This study is comprised of two phases: first, it explores the methodology of mining English learning behavior patterns based on big data, which includes how to effectively acquire, clean, integrate and analyze all kinds of data generated in the process of English learning. In the second phase, we examine how to design and implement personalized English teaching strategies based on the mined behavioral patterns [8, 9].

2. LITERATURE REVIEW

2.1 Application and Value of Big Data in Education

In recent years, the application of big data technology in the field of education has shown an increasing trend and far-reaching influence. Academic research has shown that big data is not only a source of massive information; it is also a powerful tool that can be used to reveal the deeper layers of educational phenomena and individual characteristics [10]. Specifically, the potential of big data has been fully unleashed in the niche field of English education. It can systematically capture and exhaustively record all of the students' activities in digital learning environments including, but not limited to, multidimensional data such as the selection and length of stay of reading materials, the frequency and quality of writing practice, the performance and feedback of listening training, as well as the degree of participation in online interactions and periods of online activity [11]. Through in-depth mining and intelligent analysis of these data, educators can acquire unprecedented micro-insights, pinpointing each student's learning status, skill mastery level, and potential strengths and weaknesses, so as to formulate teaching plans and tutoring strategies that are more tailored to individual characteristics. The practical application of big data technology in the field of education also facilitates the optimal organization and rational allocation of educational resources. Educational administrators are able to comprehensively grasp the student learning behavior data so as to adjust course content if required, and supply teaching resources and learning tools to ensure that educational inputs can align the service with the real needs of students and their academic progress [12]. At the same time, big data helps to refine teaching management. By means of real-time monitoring of student learning data and intelligent analysis, teachers can be apprised of the teaching and learning outcomes in real time, timely adjust schedules and methods, and effectively prevent students from struggling and feeling overwhelmed, while motivating students to persist. In short, big data technology is gradually changing the face of the education sector, to achieve truly personalized education, high-performance teaching and high-quality training of talents to provide strong technical support and data protection.

2.2 Related Research on Mining English Learning Behavior Patterns

With the advancement of information technology and data analysis tools, scholars have made significant breakthroughs in their exploration of the behavioral patterns of English learners. Leveraging the power of big data analytics tools and techniques, including but not limited to data mining algorithms, machine learning models, and complex network analysis, researchers have successfully uncovered the diverse trajectories and patterns of learning behaviors hidden behind massive amounts of learning data, especially for English language learning, which is a core subject given this era of globalization. A study by Li et al. [13] used a big data platform to track and analyze students' behavioral characteristics in digital English learning environments, and found that there are significant differences in learners' behavioral rhythms, levels of energy engagement, and their self-regulated learning behaviors (e.g., self-planning, monitoring, and adjusting the learning process) when they are learning online. These findings help educators to identify students' productive learning periods, and fluctuations in motivation, and design personalized learning plans and interventions accordingly. Taken together, these English learning behavior patterns distilled from big data analysis not only enrich our cognitive picture of the learning process, but also provide a scientific basis for educational practice [14, 15]. Based on these empirical data-driven results, educators can accurately customize teaching programs, implement tiered instruction, differentiated instruction, and intelligent instructional interventions to effectively promote each student's English learning ability and achievement. In this way, big data not only pushes the frontiers of education science research, but also promotes a new wave of reforms for education equity and quality improvement in practice.

2.3 Development and Challenges of Personalized Instructional Strategies

As an important research direction in the field of education, personalized teaching strategies have undergone a substantial transformative journey from purely theoretical conception to being driven by technological forces. Early personalized teaching concepts emphasized the crucial role of the teacher, and the development of teaching strategies relied heavily on the teacher's professionalism, experience, and unique teaching style [16]. Teachers would flexibly adjust teaching methods and content to meet students' individualized needs based on their observation, communication, and understanding of students' learning characteristics and needs. With the rapid development of big data technology and artificial intelligence, personalized teaching strategies have ushered in a new stage of development, i.e., based on data-driven customized programs [17]. This shift means that teaching

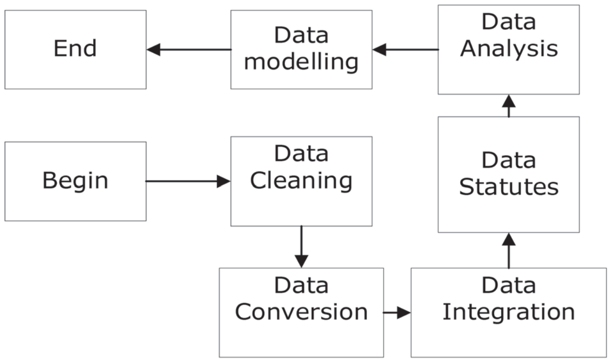

Figure 1 Data flow.

no longer relies solely on teachers' intuition and experience. Instead, it uses advanced data analytics to deeply mine the massive amount of data on students' learning behaviors, based on which personalized teaching strategies are scientifically formulated and optimized to maximally fit each student's personalized needs and learning paths.

Although big data and AI technologies offer great opportunities for personalized teaching, they still face a series of challenges in practice [18]. The first and foremost challenge is how to ensure the safety and privacy protection of students' personal information while utilizing the data, especially in a big data environment where the collection, storage and use of data need to comply with strict laws, regulations and ethical norms. Second, how to accurately predict and respond to changes in students' learning needs in a timely manner is also a major challenge, as students' learning status and needs are dynamic, not fixed, and change according to time, context, and other factors. Furthermore, how to appropriately integrate technological means into the teaching process to fully utilize the advantages of technology-assisted teaching and learning, while maintaining humanistic care and professional guidance of teachers and avoiding over-reliance on technology leading to the lack of the essence of education, is also an urgent problem to be solved. Against this background, researchers and educational practitioners work together to promote the healthy development of personalized teaching strategies through continuous exploration and refinement, striving to find a balance between respecting students' individual needs, protecting students' data privacy, making effective use of technological resources, and maintaining quality teacher guidance [19].

3. METHODOLOGY FOR MINING ENGLISH LEARNING BEHAVIOR PATTERNS BASED ON BIG DATA

3.1 Data Acquisition and Preprocessing

3.1.1 Introduction to Data Sources in the English Learning Process

The data sources examined in this study are very rich, and include:

(1) Log-in record data: containing information such as student ID, log-in date and time, log-in duration, etc.

(2) Learning activity data: such as online course viewing duration, chapter completion progress, reading material reading times and length of stay, online test scores, etc.

(3) Interactive behavior data: including the number of forum posts, discussion group participation, Q&A area question and answer records 4. Skill training data: such as the correct rate of listening training, the number of writing practice submissions, and pronunciation assessment scores [20].

3.1.2 Effective Strategies for Data Acquisition

The data can be accessed in the following ways:

(1) API interface call: through cooperation with online learning platforms, the API interface is used to obtain students' learning behavior data in real time or at regular intervals on demand. For example:

{
“student_id”: “S001”,
“timestamp”: “2022-01-01T10:00:00Z”,
“activity”: “reading”,
“duration”: 120, // in minutes
“completion_count”: 3,
“accuracy”: 0.85

}

(2) Log file analysis: user activity records are extracted from server logs, such as page visits, click events, etc. [21].

(3) Integration of third-party tools: plug-ins or sdks embedded in learning products are used to automatically collect data on user behavior.

3.1.3 Methods and Techniques for Data Cleansing and Integration

Our processing of data includes missing value processing, data consistency detection and correction, etc., as shown in Figure 1.

Missing value handling: Mean imputation is used for missing values in continuous variables, and mode imputation or nearest neighbor imputation for missing values in discrete variables [22].

Outlier detection and handling: Outliers are detected and removed using the Boxplot method with the following formula: $IQR = Q3 - Q1$, upper bound $= Q3 + 1.5 \times IQR$ lower bound $= Q1 - 1.5 \times IQR$, and observations are considered outliers if they exceed the upper and lower bounds.

Data consistency checking and correction: ensure that there are no conflicts or contradictions in the data under the same student ID in the same time interval.

Data integration: through the JOIN operation of SQL or the merge function of Python's Pandas library, data from different sources are integrated to create a unified database of student behavior [23].

3.2 Quantitative Analysis Methods of English Learning Behavior Patterns

3.2.1 Deep Mining Methods for Large-Scale Learning Behavior Data

Using advanced cluster analysis techniques, we can effectively employ the K-means algorithm for in-depth exploration and classification and generalization of student behavioral data. The method is applied to reveal the different group characteristics underlying large-scale student behavioral data, and abstract the multiple behavioral manifestations of students into multiple core behavioral patterns or types through an iterative optimization process. Firstly, K student behavior samples are randomly selected as initial clustering centers based on a preset K value. Subsequently, during each iteration, the distance between each student's behavioral indicators and each clustering center is calculated and assigned to the most similar behavioral type cluster based on the principle of minimum distance. Immediately after that, recalculate the central feature of each cluster, i.e., the statistical center of all students' behaviors in the cluster, and iterate in this way until the position of the clustering center changes tend to be stable and reach the convergence condition. K-means algorithm classifies the types of student behaviors: $argmin \sum_{i=1}^{n} \|x_i - \mu_k\|^2$. where x_i denotes the behavioral feature vector of student i, and μ_k denotes the clustering center, and the optimal clustering center is found through iterative optimization [24].

With the help of powerful pattern mining techniques, especially the prefixspan algorithm, we are able to systematically gain insights into the typical paths and habitual behavior patterns of students in the learning process. The prefixspan algorithm is applied to mine the frequent patterns in sequential data, and has been widely used in the education field to explore the learning trajectories and journeys of students. When applied to student learning behavior data, prefixspan first constructs an item set tree to record time-series data of students' learning activities, such as successive courses taken, educational resources used, and interactions engaged. The algorithm follows a depth-first search strategy to gradually expand each learning behavior segment (prefix), and determines the frequency of occurrence of each prefix pattern among all students through statistical analysis, and only prefixes that reach a preset threshold of support are retained and further explored for the possibility of their subsequent behaviors. The specific formula for the identification of student learning paths by the prefixspan algorithm is *Maximal Sequential Patterns*(MSP) $X = < x_1, x_2, \ldots, x_m >$, and under the given support threshold, to discover the longest sequence pattern that recurs during students' English learning process [25].

3.2.2 Identification and Extraction of *Significant* Learning Behavior Patterns

Association rule mining is a machine learning method for finding interesting relationships or patterns between variables in large-scale datasets, and Apriori algorithm is one of the classic applications. The Apriori algorithm is used to discover frequent itemsets and their association rules, for example, in education scenarios, by analyzing the historical learning behavior data of the students (e.g., frequency of studying course A and then course B), we can find out the correlation between courses, such as "most of the students studying course A will go on to course B", which is of great value for course recommendation and the designing of a learning path. The correlation between courses, such as "most students who study course A will continue to study course B", which is of great value for course recommendation and learning path design. Using association rule mining, such as Apriori algorithm to find out the correlation between behaviors: $support(A \Rightarrow B) \geq min\_sup$, $confidence(A \Rightarrow B) = \frac{support(A \cup B)}{support(A)} \geq min\_conf$ [26, 27].

The Autoregressive Integrated Sliding Average (ARIMA) model in time series analysis, which is a statistical model widely used for predicting future trends. When predicting students' future online learning behaviors, the ARIMA model is able to take into account factors such as trends, seasonality and random fluctuations based on the time series data of students' online learning behaviors (e.g., daily online learning hours, weekly visits to courses, etc.) over a past period of time, and then predict the learning behaviors at a certain point of time in the future. This helps education platforms to optimize services such as resource scheduling, personalized reminder push, etc., and enhance students' online learning experience and effect. The ARIMA model predicts students' future online learning behavior: $Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$, where Y_t is the value of the time series at time t, ϕ and θ are autoregressive and sliding average parameters, and ϵ is the error term [28].

3.3 Interpretation of Results and Theoretical Correlations of Behavioral Pattern Mining

As can be seen from Table 1, through K-means clustering analysis, we found three types of online learner behavioral patterns: students in Cluster 1 mainly show frequent online reading, less participation in interactions, and were especially keen on grammar practice sessions. Students in Cluster 2 show more balanced learning habits. They invested similar amounts of time in reading, listening, and writing, and were

Table 1 Cluster Analysis (K-means).

Cluster number	Cluster Center Characterization
Cluster 1	High-frequency online reading, low-frequency interaction, favoring grammar exercises
Cluster 2	Balanced reading, listening and writing, good use of fragmented time to learn
Cluster 3	Emphasis on oral communication, active participation in group discussions, preference for video tutorials

Table 2 Results of Sequence Pattern Mining.

Serial number	Longest Sequence Mode	Degree of support
Pattern 1	Watch the video → Complete the accompanying exercises → Participate in the discussion	0.35
Pattern 2	Reading articles → memorizing words → writing exercises	0.42
Pattern 3	Listening → Dictionary → Spoken Dialogue Simulation	0.28

Table 3 Results of Association Rule Mining.

Predicted behavior	Follow-up behavior	Degree of support	Confidence level (math.)
Reading materials	Complete the exercises in the back of the book	0.6	0.8
Participation in panel discussions	Timely submission of assignments	0.5	0.75
High frequency of accessing dictionary	High vocabulary test scores	0.7	0.9

Table 4 Results of Time Series Analysis

Predictive indicators	Predicted value at time $t+1$	Predicted value at time $t+2$	Time $t+3$ predicted value
Online learning hours (minutes)	65.2	66.8	67.3
Monthly Word Memorization	150	155	160
Number of tasks completed per week	10.3	10.7	11.0

good at seizing the time to improve. Cluster 3 students pay special attention to oral communication, often participate actively in group discussions, and like to use video tutorials to support their learning. As can be seen from Table 2, the results of sequence pattern mining show several common learning behavior paths: "Watching videos→completing supporting exercises→participating in discussions" is the most common behavior pattern, accounting for 35% of all behavioral sequences. The coherent learning process of "reading articles→memorizing words→writing practice" has a relatively higher level of support, reaching 42%. "Listening training → checking the dictionary → speaking simulation dialog" is also a regular combination of learning behaviors, with a support level of 28%. As can be seen in Table 3, association rule mining reveals some key behavioral associations: the data show that in 60% of the cases, students read the textbook immediately after completing the post-work problems, and the correlation between this pre- and post-behavior is very high, with a confidence level of 80%. Students who participated in group discussions were 75% likely to submit their assignments on time, showing a clear positive correlation between the two. Students who consulted the dictionary with high frequency generally performed well on the word test, a correlation rule supported by 70% with a 90% confidence level. As can be seen in Table 4, the following predictions for future trends in learning behavior are made based on time series analysis. Online learning duration is expected to reach 65.2 minutes in the next time period ($t+1$), followed by a gradual increase to 66.8 minutes ($t+2$) and 67.3 minutes ($t+3$). It is predicted that the students' monthly word memorization will continue to increase over the next few months, reaching 150 ($t+1$), 155 ($t+2$), and 160 ($t+3$), respectively. Similarly, the number of tasks completed per week will gradually increase from the current 10.3 tasks (t moments) to 10.7 ($t+1$) and 11.0 ($t+2$) [29, 30].

Tables 1–Table 4 show the results of behavioral pattern mining, according to which we can get the following six patterns

Model 1: Focused Learners and Specialized Enhancement

Interpretation of results: according to the K-means clustering analysis in Table 1, the students of Cluster 1 focused on online reading and showed a high level of enthusiasm for the grammar practice section in particular, and this mode of learning by specializing in a specific domain helped them achieve rapid progress and solid foundation building in that domain.

Model 2: Balanced learner and holistic literacy development

Interpretation of results: The Cluster 2 students in Table 1 have a more balanced input in all aspects of reading, listening and writing, and at the same time, they are good at utilizing fragmented time for self-improvement, and this all-around learning mode is conducive to the comprehensive development of all language skills and the enhancement of their comprehensive language literacy.

Model 3: Interactive learners and practical skills enhancement

Interpretation of results: the Cluster 3 students in Table 1 tended to actively participate in group discussions and use video tutorials to aid their learning, and this mode of learning, which emphasized oral communication and hands-on practice, gave them an advantage in practical application and oral communication skills.

Model 4: Organized Learning Paths and Efficient Learning Strategies

Interpretation of results: Based on the results of sequential pattern mining in Table 2, the sequential learning paths such as "watching videos→completing supporting exercises→participating in discussions", "reading articles→memorizing words→writing exercises", and "listening training→consulting the dictionary→spoken language Sequential learning paths such as "watching videos→completing supporting exercises→participating in discussions", "reading articles→memorizing words→writing exercises", and "listening training→consulting dictionaries→speaking exercises" are prevalent in the student population and have a high degree of support, and these sequential learning behavioral patterns show how students effectively organize and arrange their learning activities in order to achieve efficient learning.

Model 5: Behavioral Correlation and Optimization of Learning Outcomes

Interpretation of results: according to the association rule mining in Table 3, the strong correlation between completing post-class exercises after reading the textbook, participating in group discussions and submitting assignments on time, as well as the correlation between high-frequency dictionary consulting and excellent performance on word tests reveal the interactive effects between certain key learning behaviors, and these behavioral associations can be used as important references to optimize learning strategies.

Model 6: Predicting and guiding future learning trends

Interpretation of results: the time-series analyses in Table 4 predicted an increasing trend in online learning hours, a steady increase in monthly word recall, and an increase in the number of tasks completed per week, and these predictions can help educators anticipate and guide students' future online learning behaviors to further optimize instructional programs and individual learning plans.

4. DESIGN AND IMPLEMENTATION OF INDIVIDUALIZED INSTRUCTIONAL STRATEGIES BASED ON BEHAVIORAL MODELS

4.1 Relationship Between Behavioral Patterns and Individualized Instructional Strategies

Behavioral patterns play a key role in the design of personalized teaching strategies, which reveal students' unique habits, preferences, strengths and challenges in the process of English learning. Students' behavioral patterns obtained through big data analysis can reflect the inherent patterns and individual differences in their learning process, which provides a scientific basis for teachers to formulate precise and efficient personalized teaching strategies. For example, knowing the efficient learning behavior patterns of students in a specific period of time, teachers can arrange the key teaching activities in the corresponding period accordingly. Identifying students' patterns in the selection and use of learning resources helps teachers recommend appropriate learning materials and methods for them.

4.2 Interpretation of Behavioral Patterns From a Qualitative Research Perspective

Educational psychology theory, as an important theoretical support, provides a scientific perspective and methodological basis for deeply analyzing and guiding students' behavioral patterns. One strong example of this is self-efficacy theory. The theory emphasizes the influence of an individual's level of confidence in his or her ability to successfully perform the required behaviors to achieve the desired goals on behavioral performance. In an educational context, when educational data reveal that a group of students are significantly more effective at learning independently, it can be interpreted in light of self-efficacy theory that these students may have a greater ability to self-manage and self-regulate their learning. Accordingly, teachers can take measures to strengthen these strengths, for example, by setting up more challenging independent learning projects, so that they can continuously improve their self-driven learning abilities in practice. Cognitive load theory also plays a key role in the analysis of educational behavior patterns. This theory focuses on the impact of cognitive stress on learning effectiveness as people process information. If it is observed that students' concentration shows a significant downward trend in a specific period of time, this may be due to the excessive amount of information received in a short period of time, which leads to a high cognitive load. At this time, teachers can optimize the teaching strategy according to the cognitive load theory, adjust the difficulty and depth of the teaching content at the right time, control the pace of teaching, and avoid the cognitive overload of students triggered by the transmission of too much information at once, so as to effectively improve the teaching effect and ensure that the students efficiently absorb and master knowledge.

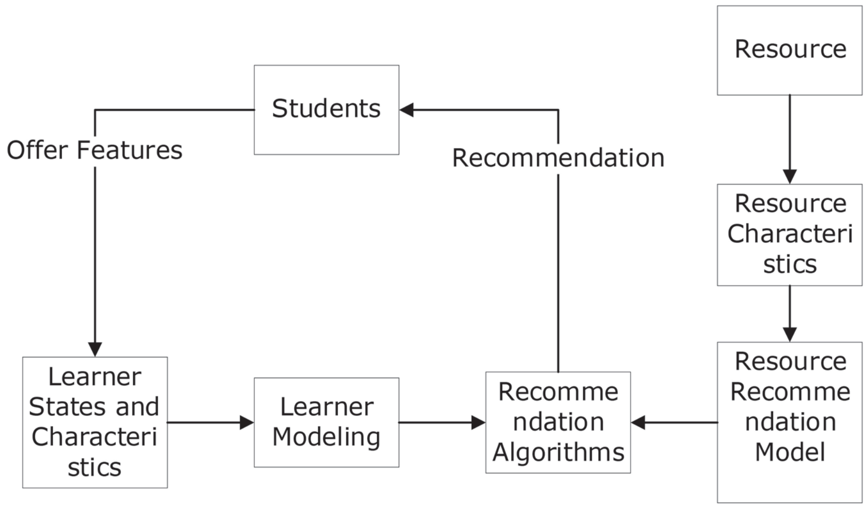

Figure 2 Individualized Teaching Model

From the perspective of constructivist learning theory, if students demonstrate positive patterns of learning behaviors in interactive exchanges and group discussions, then teachers can design more collaborative learning tasks to facilitate their construction and internalization of knowledge. Similarly, according to the deep learning theory, if students are found to demonstrate good knowledge absorption in deep reading and prolonged immersion, personalized instructional strategies should focus on providing high-quality, challenging learning resources and opportunities for deep thinking.

4.3 Design Principles and Processes of Personalized Instructional Strategies

The personalized teaching approach is shown in Figure 2. The design of personalized teaching strategies is a systematic process, and the primary principle is to focus on students' individual differences and comprehensively consider their unique learning styles, interest tendencies, and needs, on the basis of which teaching and learning activities are carried out. The design process first involves detailed data collection and in-depth mining of students' behavioral patterns, such as identifying students' learning habits, strengths and weaknesses through data analysis tools. Subsequently, the mined behavioral patterns are combined with education and teaching theories to accurately interpret students' personalized learning characteristics. Based on this, an accurate analysis of teaching needs is carried out, and teaching goals are then set to meet students' personalized development needs. In regard to goal-setting, teachers need to flexibly utilize a range of teaching methods and technologies, such as project-based learning, microclasses, blended learning, etc., to establish diversified learning environments that can be adapted to the needs of different students. At the same time, they should ensure that the teaching content, methods and evaluation system are personalized as much as is practical. In the implementation stage, teachers teach according to the established teaching strategies and simultaneously carry out real-time monitoring and evaluation of students' learning outcomes.

4.4 Specific Implementation Plan for Personalized Instructional Strategies

Taking students with high online learning efficiency during evening self-study as an example, according to the theory of biological rhythms, everyone has his or her own efficient learning time period, and some students are able to concentrate more in the evening, so special arrangements for online Q&A and discussion slots in the evening can align with the physiological characteristics of such students and maximize the efficiency of their learning window, thus improving learning efficiency. The following specific strategies can be adopted:

(1) According to the theory of biological rhythms, special online Q&A time and discussion sessions are arranged in the evening to make full and efficient use of students' study time.

(2) Combined with data analysis of the learning platform, students are provided with personalized learning task packages that include English reading materials, listening training and vocabulary review resources that match their learning habits and pace.

(3) The establishment of a goal-oriented independent learning program to encourage students to make use of their evening study time for in-depth learning and independent exploration, is complemented by appropriate teacher guidance and support.

(4) Through the Intelligent Learning System (ILS), students' learning behaviors during evening self-study are monitored in real time, and timely feedback and suggestions are provided to support the continuous improvement of their learning strategies.

In conclusion, the design and implementation of personalized teaching strategies based on behavioral patterns is a cyclical and dynamic adjustment process, which needs to be closely combined with the results of big data analysis and the theories of pedagogy and psychology, in order to teach

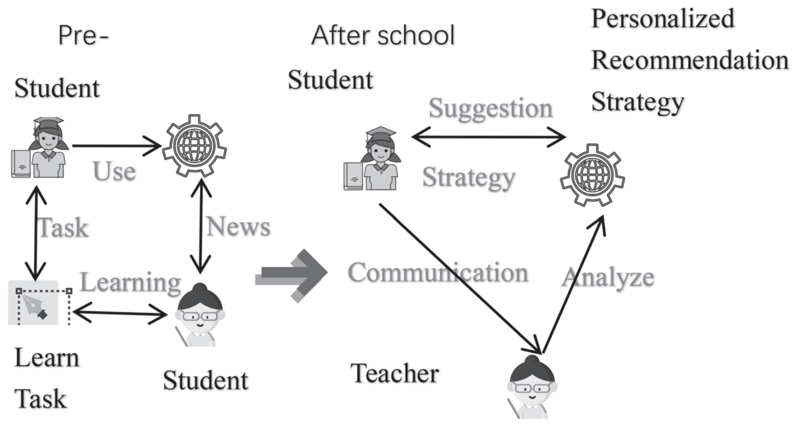

Figure 3 Learning Pattern of Students in the Experimental Group.

Table 5 Basic Information of Experimental and Control Groups.

Groups	Number of students (N)	Average age (Age)	Male to female ratio (M:F ratio)	Initial English Proficiency (Level)
Experimental group	100	20 years old	1:1.5	A1
Control subjects	100	20 years old	1:1.5	A2

according to the students' abilities and improve the quality and effectiveness of English teaching.

5. EXPERIMENTAL VALIDATION AND CASE STUDY OF PERSONALIZED TEACHING STRATEGIES

We selected a certain number of English learners and classified them according to the behavioral patterns found in the preliminary big data analysis, with each category representing a typical learning behavioral pattern. Then, we randomly divided the participants into an experimental group (implementing personalized teaching strategies) and a control group (traditional teaching mode) to ensure that the two groups are balanced in terms of age, gender, and initial English level. Further, based on the results of the behavioral pattern analysis, individualized teaching strategies were designed and implemented for the experimental group, while the control group were exposed to the conventional teaching methods. Two types of indicator data were collected before and after the experiment: first, learning behavior data, including online learning hours, number of interactions, and task completion. The second is academic performance data, including test scores, homework grading, etc. The details of the strategy implementation process, such as student feedback and teacher observation, are also recorded. The learning pattern of students in the experimental group is shown in Figure 3.

Table 5 clearly shows the balance between the experimental and control groups in terms of the number of students, age, gender ratio, and initial English proficiency. This helps to ensure the accuracy of the experimental results because the consistency of these variables between the two groups reduces possible confounders, allowing any observed differences to be attributed to differences in teaching strategies.

Table 6 shows that the experimental group's online learning hours, number of interactions and task completion were significantly improved after the implementation of personalized teaching strategies. Compared with the control group, the improvement of these indicators is more significant, suggesting that personalized teaching strategies can effectively improve students' learning behaviors.

Figure 4 shows that the experimental group's academic performance increased significantly in terms of final exam results and homework scores. Although the control group also improved, the experimental group's improvement was greater, further validating the effectiveness of the individualized teaching strategy.

By analyzing students' behavioral patterns and personalized strategies, this study aims to identify students who can learn more efficiently at night so that they can adjust their study plans and optimize their personal learning outcomes. For example, after student ID1 was identified as an efficient learner in the evening, an evening online Q&A service was customized for him. The results showed that his evening learning efficiency increased by 30% and his final grade increased by 15%. Student ID2, as a student with strong self-learning ability, achieved a 20% faster task completion speed and mastered new knowledge faster by pushing high-level self-learning tasks. The implementation of these personalized strategies significantly improved students' learning efficiency and grades.

Table 7 demonstrates, through relevant cases, how personalized teaching strategies can produce positive effects according to the different characteristics and needs of students. The significant improvement in learning efficiency and achievement of ID1 and ID2 highlights the flexibility and practicality of personalized teaching strategies.

Table 6 Comparison of Learning Behavior Data before and after the Experiment.

Behavioral indicators	Experimental group pre-test (T1_E)	Experimental group post-test (T2_E)	Control group pre-test (T1_C)	Control group post-test (T2_C)
Hours of online learning (hours)	5	8	4	6
Number of interactions	20	30	15	22
Degree of mission accomplishment	80%	95%	75%	85%

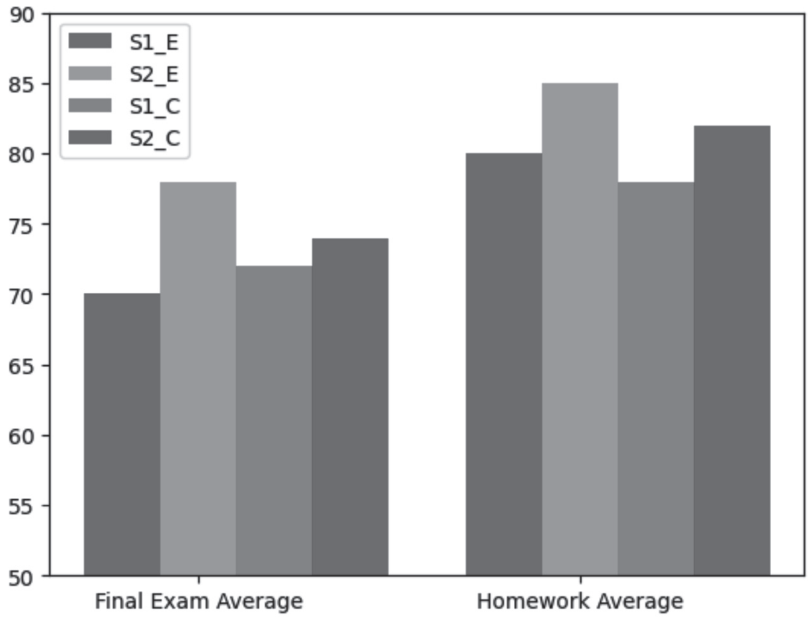

Figure 4 Comparison of Academic Performance before and after the Experiment.

Table 7 Typical Examples of Individualized Instructional Strategies.

Student ID	Behavioral model	Personalization Strategies	Post-implementation changes	Student ID
ID1	Evening Highly Effective Learners	Customized evening online Q&A service	30% increase in evening study efficiency and 15% increase in final grades	ID1
ID2	Self-directed learning powerhouse	Push high-level self-directed learning tasks	20% faster task completion and quicker acquisition of new knowledge	ID2

Table 8 Summary of Strategy Implementation Issues and Solutions.

Concern	Prescription
Individual students are not receptive to new technology	Provide customized training and additional technical support to gradually familiarize students with new technologies
Data Privacy Protection Concerns	Strengthen cybersecurity measures, clarify privacy protection policies and user agreements to ensure compliant operations
Difficulty for teachers in adapting to new teaching models	Conduct a series of professional trainings and seminars, share successful cases, and provide regular feedback and strategy adjustments

Table 8 lists the solutions to problems that may be encountered during the implementation of individualized instructional strategies. This demonstrates the program team's anticipation of potential challenges when planning and implementing the strategies and preparing strategies to address them.

The experimental data and case studies suggest that personalized teaching strategies are an effective means of improving students' learning behaviors and academic performance. The comparison between the experimental and control groups, as well as the results of individual cases, point to the advantages of personalized teaching strategies. Of course, the achievement of these results relies on careful strategy design, effective implementation, and rigorous monitoring of the process and results. In the future, personalized instructional strategies can be further optimized through continuous data collection and analysis to accommodate changing educational needs and student diversity.

6. CONCLUSION

In this paper, through the big data mining and analysis process, we constructed a complete set of students' English learning behavioral patterns, and with the help of quantitative methods such as clustering analysis and association rule analysis, we conducted an in-depth exploration of the characteristics of students' behaviors when learning English. We found that behavioral patterns play a decisive role in the design of personalized teaching strategies, which reveal students' learning habits, interest biases, areas of strength and difficulties faced by individual students, enabling teachers to formulate precise and efficient teaching plans based on scientific evidence. The experimental data and the case study together verified the significant effect of personalized teaching strategies on improving students' learning behavior and academic achievement. By comparing the differences in performance between the experimental group and the control group, as well as the in-depth analysis of individual cases, we highlight the unique advantages of individualized teaching strategies. However, the effective implementation of individualized teaching strategies requires rigorous design, detailed implementation, and close monitoring of the teaching process and outcomes. The significance of this study lies in the fact that, for the first time, we have scientifically quantified and analyzed the English learning behavior patterns through big data analysis technology, and successfully integrated them into the construction and implementation of personalized teaching strategies, thus advancing the practice of targeted and personalized teaching in the field of education.

REFERENCES

1. Ashwin, T. S., & Guddeti, R. M. R. (2019). Unobtrusive behavioral analysis of students in classroom environment using Non-Verbal cues. IEEE Access, 7, 150693–150709. doi:10.1109/access.2019.2947519
2. Cagliero, L., Farinetti, L., & Baralis, E. (2019). Recommending personalized summaries of teaching materials. IEEE Access, 7, 22729–22739. doi:10.1109/access.2019.2899655
3. Zhang, T. (2024). Design of virtual reality visualization display system based on big data technology. Engineering Intelligent Systems, 32(3), 225–233.
4. Chen, H. M., Nguyen, B. A., Dow, C. R., Hsueh, N. L., & Liu, A. C. (2022). Exploring time-related micro-behavioral patterns in a python programming online course. Journal of Information Science and Engineering, 38(6), 1109–1131. doi:10.6688/jise.202211_38(6).0002
5. Chen, J. Y., Zhou, G. Z., & Zhong, Y. Z. (2024). A personalized recommendation system for teaching resources in sports using fuzzy <i>C</i>-means clustering technique. Soft Computing, 28(1), 703–720. doi:10.1007/s00500-023-09421-0
6. Sun, X. (2024). A clustering analysis of students' English scores after targeted improvement. *Engineering Intelligent Systems,* 32(2), 89–93.
7. Chen, W. Q., & Yang, T. (2023). A recommendation system of personalized resource reliability for online teaching system under large-scale user access. Mobile Networks & Applications, 12. doi:10.1007/s11036-023-02194-8
8. Chen, Y. Y., Wang, X., & Du, X. H. (2021). Diagnostic evaluation model of English learning based on machine learning. Journal of Intelligent & Fuzzy Systems, 40(2), 2169–2179. doi:10.3233/jifs-189216
9. Fan, W. J. (2019). Influence of digital learning environment on students' English learning satisfaction and learning engagement. Journal of Environmental Protection and Ecology, 20, 281–285. Retrieved from <Go to ISI>://WOS:000531884000035
10. Herrmann, B., Araz, K., & Johnsrude, I. S. (2021). Sustained neural activity correlates with rapid perceptual learning of auditory patterns. Neuroimage, 238, 11. doi:10.1016/j.neuroimage.2021.118238
11. Hou, X. S., Fang, D., & Guo, J. (2024). Research on personalized path recommendation of college physical education online teaching based on improved particle swarm optimization. Revista Internacional De Medicina Y Ciencias De La Actividad Fisica Y Del Deporte, 24(94), 119–135. doi:10.15366/rimcafd2024.94.009
12. Khan, M. F., Lu, L., Afzal, H., & Nasir, J. A. (2023). PredictionMiner: mining the latest individual behavioral rules for personalized contextual pattern predictions. Soft Computing, 15. doi:10.1007/s00500-023-08572-4
13. Li, F., Lu, Y., Ma, Q., Gao, J. T., Wang, Z. B., & Bai, L. (2023). SPOC online video learning clustering analysis: Identifying learners' group behavior characteristics. Computer Applications in Engineering Education, 31(4), 1059–1077. doi:10.1002/cae.22624
14. Li, M. (2018). A study on the influence of non-intelligence factors on College Students' English learning achievement based on C4.5 algorithm of decision tree. Wireless Personal Communications, 102(2), 1213–1222. doi:10.1007/s11277-017-5177-0
15. Liu, H. G., Li, J., & Fang, F. (2022). Examining the complexity between boredom and engagement in English learning: evidence from Chinese high school students. Sustainability, 14(24), 12. doi:10.3390/su142416920
16. Liu, M. H. (2023). English classroom anxiety, learning style and English achievement in Chinese university EFL students. Sustainability, 15(18), 14. doi:10.3390/su151813697
17. Metruk, R. (2021). The use of smartphone English language learning apps in the process of learning English: Slovak EFL students' perspectives. Sustainability, 13(15), 17. doi:10.3390/su13158205
18. Mourtzis, D., Panopoulos, N., & Angelopoulos, J. (2023). A hybrid teaching factory model towards personalized education 4.0. International Journal of Computer Integrated Manufacturing, 36(12), 1739–1759. doi:10.1080/0951192x.2022.2145025
19. Mulualem, Y. G., Mulu, Y. E., & Gebremeskal, T. G. (2022). Effects of English learning beliefs on English achievement: academic emotions as mediators. Heliyon, 8(7), 14. doi:10.1016/j.heliyon.2022.e09829

20. Mulualem, Y. G., Mulu, Y. E., & Gebremeskal, T. G. (2022). Relations of previous English achievement with college academic achievements as mediated by English learning beliefs. Heliyon, 8(7), 9. doi:10.1016/j.heliyon.2022.e09841
21. Peng, N. F. (2022). Research on the effectiveness of English online learning based on neural network. Neural Computing & Applications, 34(4), 2543–2554. doi:10.1007/s00521-021-05855-5
22. Sáiz-Manzanares, M. C., Alonso-Martínez, L., Rodríguez, A. C., & Martin, C. (2022). Project-Based learning guidelines for health sciences students: an analysis with data mining and qualitative techniques. Jove-Journal of Visualized Experiments (190), 20. doi:10.3791/63601
23. Sergio, W. L., Ströele, V., Dantas, M., Braga, R., & Macedo, D. D. (2024). Enhancing well-being in modern education: a comprehensive eHealth proposal for managing stress and anxiety based on machine learning. Internet of Things, 25, 24. doi:10.1016/j.iot.2023.101055
24. Wang, C. X., Fang, T., & Gu, Y. X. (2020). Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia. Computers & Education, 143, 14. doi:10.1016/j.compedu.2019.103683
25. Wu, X. Y. (2018). Study on English learning system based on the difficulty of English texts. Wireless Personal Communications, 102(4), 3557–3571. doi:10.1007/s11277-018-5391-4
26. Xu, L. L., Wang, H. H., Chen, J. X., Zhang, Y. W., Huang, Z. Q., & Yu, C. F. (2022). English learning stress, self-efficacy, and burnout among undergraduate students: the moderating effect of mindfulness and gender. International Journal of Environmental Research and Public Health, 19(23), 13. doi:10.3390/ijerph192315819
27. Ye, Y. S. (2022). An analysis of the impact of college English education on students' learning anxiety. Psychiatria Danubina, 34, S696-S701. Retrieved from <Go to ISI>://WOS:000807389000444
28. Yu, J. (2020). Analysis of task degree of English learning based on deep learning framework and image target recognition. Journal of Intelligent & Fuzzy Systems, 39(2), 1903–1914. doi:10.3233/jifs-179961
29. Zarzour, H., Bendjaballah, S., & Harirche, H. (2020). Exploring the behavioral patterns of students learning with a Facebook-based e-book approach. Computers & Education, 156, 9. doi:10.1016/j.compedu.2020.103957
30. Zhang, G. J. (2022). A study on alleviating emotional anxiety of business english majors in language learning under educational psychology. Psychiatria Danubina, 34, S1005-S1010. Retrieved from <Go to ISI>://WOS:000830830000650